

The opinion in support of the decision being entered today was not written for publication and is not binding precedent of the board

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE BOARD OF PATENT APPEALS  
AND INTERFERENCES

---

*Ex parte* JEFFREY THOMAS KREULEN  
and WILLIAM SCOTT SPANGLER

---

Appeal No. 2007-0429  
Application 09/848,430<sup>1</sup>  
Technology Center 2100

---

Decided: June 8, 2007

---

Before JAMES D. THOMAS, JOHN C. MARTIN, and JAY P. LUCAS,  
*Administrative Patent Judges.*

MARTIN, *Administrative Patent Judge.*

**DECISION ON APPEAL**

This is an appeal from the Examiner's rejection of claims 1, 2, 5, 6, 9, 10, and 13-25 under 35 U.S.C. § 103(a).<sup>2</sup>

---

<sup>1</sup> Filed May 4, 2001.

<sup>2</sup> At page 13 of the Answer, the Examiner withdrew the § 103(a) rejection of claims 3, 4, 7, 8, 11, and 12. The appeal is therefore dismissed with respect to those claims.

We have jurisdiction under 35 U.S.C. §§ 6(b) and 134(a). We reverse and enter a new ground of rejection.

### **APPELLANTS' INVENTION**

Appellants' invention is a method of converting a document corpus containing an ordered plurality of documents into a compact form for storage in a memory. Figure 1 shows an example of a document corpus consisting of three documents.

#### **FIGURE 1 DOCUMENT CORPUS**

DOC1: Hello, my name is Fred.  
DOC2: Hello, my name is Scott.  
DOC3: Scott says, "Live and let live."

Appellants' method involves assigning to the significant terms "name," "fred," "scott," and "live" the integer numbers 1-4, respectively, as depicted in Figure 2:

#### **FIGURE 2 DICTIONARY**

1. name
2. fred
3. scott
4. Live

The integers are then used to represent the document corpus as a “small sparse matrix in vector form,” which is depicted as follows in Figure 7:

**FIGURE 7 SMALL SPARSE MATRIX IN VECTOR FORM**

ALLDATA = 1 2 1 3 3 4 4  
STARTMARKER = 1,3,5  
MULT = 0.707 0.707 0.447

This figure shows three vectors. The “ALLDATA” vector represents the order of appearance of the terms in the document corpus, i.e., “name,” “fred,” “name,” “scott,” “scott,” “live,” and “live” (Specification 9:12-15). The “STARTMARKER” vector represents the starting point of each document in the ALLDATA sequence (*id.* at 9:15-16). The STARTMARKER array in this example indicates that the first terms of Documents 1, 2, and 3 correspond to the first, third and fifth integers in the ALDATA array. The “MULT” vector consists of normalization factors to be applied to the terms in Documents 1-3, respectively, when floating point representation is required (*id.* at 9:16-18).

### THE CLAIMS

The independent claims on appeal are claims 1, 5, 9, 13, and 15, of which claim 1 reads:<sup>3</sup>

1. A method of converting a document corpus containing an ordered plurality of documents into a compact representation in memory of occurrence data, said method comprising:

---

<sup>3</sup> A corrected Claims Appendix was filed on June 14, 2006.

developing a first vector for said entire corpus, said first vector being a listing of integers corresponding to terms in said documents such that each said document in said document corpus is sequentially represented in said listing.

### THE REFERENCES AND REJECTIONS

The rejections are based on the following references:

Pirolli	US 5,895,470	Apr. 20, 1999
Cohen	US 5,950,189	Sep. 7, 1999
Call	US 2002/0165707 A1	Nov. 7, 2002 (filed Feb. 26, 2001)

Claims 1, 5, 9, 13, 15, and 17-25 stand rejected under 35 U.S.C. § 103(a) for obviousness over Pirolli in view of Call (Final Office Action 3).<sup>4</sup> Of the claims rejected on this ground, Appellants specifically argue the language of only claim 1 (*see, e.g.*, Br. 5). We likewise will limit our consideration of this ground of rejection to that claim. 37 C.F.R. § 41.37(c)(1)(vii)(2005).

Dependent claims 2, 6, 10, 14, and 16 stand rejected under § 103(a) for obviousness over Pirolli in view of Call and further in view of Cohen (Final Office Action 7; Answer 9). Appellants argue these claims as a group. We will limit our consideration of this ground of rejection to claim 2.

---

<sup>4</sup> The statement of this rejection in the Answer at page 4 incorrectly fails to mention claims 17-25, although they are addressed in the discussion of the rejection (Continued on next page.)

## THE ISSUE

Have Appellants demonstrated that the Examiner failed to establish the obviousness of the subject matter recited in claims 1 and 2?<sup>5</sup>

## PRINCIPLES OF LAW

“[T]he examiner bears the initial burden, on review of the prior art or on any other ground, of presenting a prima facie case of unpatentability.” *In re Oetiker*, 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992). A rejection under 35 U.S.C. § 103(a) must be based on the factual determinations required by the, namely, (1) the scope and content of the prior art, (2) the level of ordinary skill in the art, (3) the differences between the claimed invention and the prior art, and (4) objective indicia of non-obviousness. *DyStar Textilfarben GmbH & Co. Deutschland KG v. C.H. Patrick Co.*, 464 F.3d 1356, 1360, 80 USPQ2d 1641, 1645 (Fed. Cir. 2006) (citing *Graham v. John Deere Co.*, 383 U.S. 1, 17, 148 USPQ 459, 467 (1966)). In addition, “there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.” *KSR Int’l v. Teleflex, Inc.*, 127 S. Ct. 1727, 1741, 82 USPQ2d 1385, 1396 (2007) (quoting *In re Kahn*, 441 F.3d 977, 987, 78 USPQ2d 1329, 1336 (Fed. Cir. 2006)). Such

---

(*id.* at 7-9) .

<sup>5</sup> As explained in *In re Rouffet*, 149 F.3d 1350, 1355, 47 USPQ2d 1453, 1455 (Fed. Cir. 1998): “On appeal to the Board, an applicant can overcome a rejection [for obviousness] by showing insufficient evidence of *prima facie* obviousness or by rebutting the *prima facie* case with evidence of secondary indicia of nonobviousness.”

reasoning can be based on interrelated teachings of multiple patents, the effects of demands known to the design community or present in the marketplace, and the background knowledge possessed by a person having ordinary skill in the art. *KSR*, 127 S. Ct. at 1731, 82 USPQ2d at 1396.

#### THE REJECTION BASED ON PIROLI AND CALL

We begin our analysis with Call, which discloses converting natural language text data into a series of integers in order to reduce storage space requirements and permit more rapid processing than can be accomplished using character data (Call para. 9).

Referring to Figure 1, incoming text data in “a conventional file, string or character array as seen at 111” is parsed (at 115) into substrings consisting of alphanumeric terms and punctuation terms 117 (*id.* para. 26). These terms are compared (at 121) to the terms already stored in stored in term table 125 (*id.* paras. 26-27). If there is a match, the integer that has already been assigned to that term is appended to the end of integer array 135 (*id.* para. 29). If there is no match, the term is added to the term table as shown at 137 and a newly assigned integer is appended (at 136) to the end of the integer array (*id.*). At the conclusion of the parsing and storage process, the term table holds one copy of each unique substring (both alphanumeric terms and punctuation terms) that appeared in the original text data and the integer array holds an integer representing each term produced from the parser in the order found, with one exception, which is that single spaces between alphanumeric spaces are not represented (*id.* para. 30). The Examiner is

therefore correct to characterize Call as developing an “uninterrupted array of integers corresponding to an occurrence of terms” (Answer), a characterization that reflects the language employed in claim 5.<sup>6</sup> The Examiner did not find that Call discloses or suggests using the disclosed method to develop an array of integers that represents a plurality of documents.

As evidence that it would have been obvious to use Call’s method to generate a listing of integers representing a document corpus that contains an ordered plurality of documents, the Examiner relies on Pirolli. Pirolli discloses a system for analyzing the topology, content, and usage of collections of linked documents such as those found on the World Wide Web (hereinafter, Web) in order to facilitate information searching or improving the design of a web locality (col. 1, l. 65 to col. 2, l. 1). A web locality is as “[a] collection of related web pages associated with an entity having a site on the World-Wide Web such as a company, educational institute or the like” (col. 3, ll. 41-43). Pirolli’s system provides (a) categorization based on “feature vectors” that characterize individual page information and (b) prediction of the need for (or relevance of) other Web pages with respect to a particular context, which could be a particular page or set of pages,

---

<sup>6</sup> Claim 5, for example, reads:

5. A method of converting, organizing, and representing in a computer memory a document corpus containing an ordered plurality of documents, said method comprising:

for said document corpus, taking in sequence each said ordered document and developing a first uninterrupted listing of integers to correspond to an occurrence of terms in the document corpus.

using a spreading activation technique (col. 2, ll. 3-8). The Examiner held that the pages which form a website can accurately be characterized as a document corpus containing an ordered plurality of documents because “[a]s is known in the art, a website contains a home page, which frequently serves as a table of contents, with links to various subsequent pages based upon content (See Microsoft Computer Dictionary, Third Edition, page 506, definition of ‘website’, Exhibit B attached).” (Answer 11.) Appellants disagree (Reply Br. 2<sup>7</sup>). However, assuming for the sake of argument that the Examiner is correct on this point, the rejection nevertheless fails because, as pointed out by Appellants, Pirolli fails to disclose or suggest using a single vector to represent a plurality of web pages (Br. 13). We therefore agree with Appellants that even assuming it is proper to combine the teachings of Call and Pirolli,

the most that can reasonably be asserted is that Call would provide to Pirolli a preliminary conversion of data in the Web pages to be in integer format. Each page, however, remains as an isolated entity, so that the Web site (document corpus) remains as a collection of documents represented in an integer format, to now be processed as separate documents in accordance with the method described in Pirolli that includes developing a matrix, similar to the conventional methods disclosed by Appellants in their background discussion.

*Id.* We note that the Answer does not contain a response to this argument.

Because Appellants have persuaded us that Call and Pirolli considered in combination fail to disclose or suggest using a listing of integers to represent plural

---

<sup>7</sup> The title of the Reply Brief is “Appellants’ Response to Examiner’s Answer.”



documents in a document corpus, as required by claim 1, we are reversing the § 103(a) rejection for obviousness over Call in view of Pirolli with respect to that claim as well as with respect to the other claims grouped therewith, namely, independent claims 5, 9, 13, and 15 and dependent claims 17-25.

#### THE REJECTION BASED ON PIROLI, CALL, AND COHEN

The § 103(a) rejection of dependent claim 2 and of the other claims grouped therewith, namely, claims 6, 10, 14, and 16, is reversed because the above-noted deficiency in Call and Pirolli is not cured by the subject matter the Examiner relies on in Cohen. Cohen is cited by the Examiner only for a teaching of developing a normalized vector containing floating point multipliers (Answer 9).

#### NEW GROUND OF REJECTION

Pursuant to our authority under 37 C.F.R. § 41.50(b) (2006), we are entering a new ground of rejection of independent claims 1, 5, 9, 13, and 15 and dependent claims 18 and 20-22 under 35 U.S.C. § 102(e) for anticipation by Call. Call discloses that “an e-book player using the present invention could store a book library four times as large as would be possible using indexed text files, while providing more rapid and more robust search and display capabilities” (Call para. 139). We understand this passage to mean that each e-book is converted to and stored as a respective integer file. For the following reasons, we hold that phrase “document corpus containing an ordered plurality of documents,” which appears in each of the independent claims, reads on one of the books in Call’s e-

book library. We begin by take Official Notice that many books include a plurality of chapters and that some of the books in Call's e-book library necessarily will contain plural chapters. *See In re Ahlert*, 424 F.2d 1088, 1091, 165 USPQ 418, 420 (CCPA 1970) (PTO tribunals are permitted to take notice of facts beyond the record which, while not generally notorious, are capable of such instant and unquestionable demonstration as to defy dispute) (citing *In re Knapp Monarch Co.*, 296 F.2d 230, 132 USPQ 6 (CCPA 1961)).

Next, we note that the term "document" is not defined in Appellants' Specification and therefore must be given its broadest reasonable interpretation consistent with Appellants' disclosure. *In re Bigio*, 381 F.3d 1320, 1325, 72 USPQ2d 1209, 1211 (Fed. Cir. 2004). A definition of this term has not been provided by the Examiner or Appellants. The following definitions are relevant:

**document** (1) A named, structural unit of text that can be stored, retrieved, and exchanged among systems and users as a separate unit. (T)<sup>8</sup> (2) Information and the medium on which it is recorded that generally have permanence and can be read by humans or by machine. (3) In word processing, a collection of information that pertains to a particular subject or related subjects. Synonymous with file. (4) In Systems/36, a collection of one or more lines of text that can be named and stored as a member in a folder.

---

<sup>8</sup> The letter "T" indicates that the definition is taken from draft international standards, committee drafts, and working papers being developed by the ISO/IEC/JTC1/SC1 and indicates that final agreement has not been reached among the participating National Bodies of SC1. George McDaniel, *IBM Dictionary of Computing* vii, viii, 212 (10<sup>th</sup> Ed. Aug. 1993) ("IBM Dictionary") (copy enclosed).

*IBM Dictionary 212* (copy enclosed). Definition (1) is broad enough, in our view, to read on a chapter of an e-book that contains plural chapters, with the result that such an e-book constitutes a “document corpus containing an ordered plurality of documents” in the sense of claims 1, 5, 9, 13, and 15. The result of applying Call’s conversion method to such an e-book is to generate a first vector in the form of a first uninterrupted listing of integers corresponding to terms in the documents (i.e., chapters) such that each said document in the document corpus (i.e., e-book) is sequentially represented in the first vector, thereby satisfying each of the independent claims.

Dependent claims 18 and 20-22, which call for developing a dictionary comprising the terms contained in the document corpus, read on Call’s display term list 152 in Figure 1 (Call para. 32).

The remaining dependent claims recite -- or depend on claims that recite -- features not disclosed by Carr, including (a) developing a second vector or listing indicating the location of each document’s representation in the first vector or first listing (claims 19, 21, 23, and 25); (b) rearranging the order of the integers within the data for each document so that all identical unique integers are adjacent (claims 3, 7, 11, and 17); and (c) developing a third vector or listing containing floating point multipliers (claims 2, 6, 10, 14, and 16). Pursuant to the discretion accorded the Board by 37 C.F.R. § 41.50(b), we have limited our consideration of new grounds of rejection to the issue of anticipation by Carr.

### DECISION

The rejection of claims 1, 5, 9, 13, 15, and 17-25 under § 103(a) for obviousness over Pirolli in view of Call is reversed, as is the rejection of claims 2, 6, 10, 14, and 16 under § 103(a) for obviousness over Pirolli in view of Call and further in view of Cohen.

Furthermore, pursuant to our authority under 37 C.F.R. § 41.50(b) (2006), we have entered a new ground of rejection of claims 1, 5, 9, 13, 15, 18, and 20-22 under 35 U.S.C. § 102(e) for anticipation by Call.

### APELLANTS' OPTIONS FOR RESPONDING TO THE NEW GROUND OF REJECTION

Appellants' options for responding to the new ground of rejection are as follows:

(b) . . . A new ground of rejection pursuant to this paragraph shall not be considered final for judicial review. When the Board makes a new ground of rejection, the appellant, within **two months** from the date of the decision, must exercise one of the following two options with respect to the new ground of rejection to avoid termination of the appeal as to the rejected claims:

(1) *Reopen prosecution*. Submit an appropriate amendment of the claims so rejected or new evidence relating to the claims so rejected, or both, and have the matter reconsidered by the examiner, in which event the proceeding will be remanded to the examiner. The new ground of rejection is binding upon the examiner unless an amendment or new evidence not previously of record is made which, in the opinion of the examiner, overcomes the new ground of rejection stated

Appeal No. 2007-0429  
Application 09/848,430

in the decision. Should the examiner reject the claims, appellant may again appeal to the Board pursuant to this subpart.

(2) *Request rehearing.* Request that the proceeding be reheard under § 41.52 by the Board upon the same record. The request for rehearing must address any new ground of rejection and state with particularity the points believed to have been misapprehended or overlooked in entering the new ground of rejection and also state all other grounds upon which rehearing is sought.

37 CFR § 41.50(b) (2006) (bolding added.)

No time period for taking any subsequent action in connection with this appeal may be extended under 37 CFR § 1.136(a). *See* 37 CFR §§ 41.50(f) and 41.52(b) (2006).

REVERSED; 37 C.F.R. § 41.50(b)

pgc

MCGINN INTELLECTUAL PROPERTY LAW GROUP, PLLC  
8321 OLD COURTHOUSE ROAD  
SUITE 200  
VIENNA, VA 22182-3817

Enclosure: George McDaniel, *IBM Dictionary of Computing*, vii, viii, 212  
(10<sup>th</sup> Ed. Aug. 1993).